

APPROXIMATE CENTRAL LIMIT THEOREMS

BEN BERCKMOES AND GEERT MOLENBERGHS

ABSTRACT. We refine the classical Lindeberg-Feller central limit theorem by obtaining asymptotic bounds on the Kolmogorov distance, the Wasserstein distance, and the parametrized Prokhorov distances in terms of a Lindeberg index. We thus obtain more general approximate central limit theorems, which roughly state that the row-wise sums of a triangular array are approximately asymptotically normal if the array approximately satisfies Lindeberg's condition. This allows us to continue to provide information in non-standard settings in which the classical central limit theorem fails to hold. Stein's method plays a key role in the development of this theory.

1. INTRODUCTION

Throughout, we assume that all random variables are defined on a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Let ξ be a standard normal random variable, that is, a normally distributed random variable with $\mathbb{E}[\xi] = 0$ and $\mathbb{E}[\xi^2] = 1$, and $\{\xi_{n,k}\}$ a standard triangular array (STA) of random variables, that is, a triangular array

$$\begin{array}{ccccc} & & \xi_{1,1} & & \\ & & \xi_{2,1} & \xi_{2,2} & \\ & & \xi_{3,1} & \xi_{3,2} & \xi_{3,3} \\ & & & \vdots & \end{array}$$

of random variables with $\xi_{n,1}, \dots, \xi_{n,n}$ independent for all n , $\mathbb{E}[\xi_{n,k}] = 0$ for all n, k , and $\sum_{k=1}^n \mathbb{E}[\xi_{n,k}^2] = 1$ for all n .

Recall that the sequence $(\sum_{k=1}^n \xi_{n,k})_n$ is said to converge weakly to ξ iff

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\sum_{k=1}^n \xi_{n,k} \leq x_0 \right] = \mathbb{P}[\xi \leq x_0]$$

Ben Berckmoes is post doctoral fellow at the Fund for Scientific Research of Flanders (FWO).

Geert Molenberghs gratefully acknowledges financial support from the IAP research network #P7/06 of the Belgian Government (Belgian Science Policy).

for all x_0 at which the map $x \mapsto \mathbb{P}[\xi \leq x]$ is continuous, or, equivalently, iff

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[h \left(\sum_{k=1}^n \xi_{n,k} \right) \right] = \mathbb{E}[h(\xi)]$$

for all $h : \mathbb{R} \rightarrow \mathbb{R}$ bounded and continuous.

We say that $\{\xi_{n,k}\}$ satisfies Feller's condition iff

$$\lim_{n \rightarrow \infty} \max_{k=1}^n \mathbb{E} [\xi_{n,k}^2] = 0, \quad (1)$$

and Lindeberg's condition iff

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{E} [\xi_{n,k}^2; |\xi_{n,k}| > \epsilon] = 0$$

for all $\epsilon > 0$. It is easily seen that Lindeberg's condition implies Feller's, but that the converse does not hold.

The above language allows us to formulate the following result, which belongs to the heart of classical probability theory.

Theorem 1.1 (Lindeberg-Feller Central Limit Theorem). *Let ξ and $\{\xi_{n,k}\}$ be as above. If $\{\xi_{n,k}\}$ satisfies Lindeberg's condition, then the sequence $(\sum_{k=1}^n \xi_{n,k})_n$ converges weakly to ξ . The converse holds if $\{\xi_{n,k}\}$ satisfies Feller's condition.*

The number

$$\text{Lin}(\{\xi_{n,k}\}) = \sup_{\epsilon > 0} \limsup_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{E} [\xi_{n,k}^2; |\xi_{n,k}| > \epsilon] \quad (2)$$

was introduced in [BLV13] as the Lindeberg index. Notice that it produces for each STA a number between 0 and 1, and that it is 0 if and only if Lindeberg's condition is satisfied. So it can be thought of as a number which measures how far a given STA deviates from satisfying Lindeberg's condition.

Furthermore, let $d(\eta, \eta')$ be a metric on random variables with the property that $\lim_{n \rightarrow \infty} d(\eta, \eta_n) = 0$ is equivalent with weak convergence of $(\eta_n)_n$ to η , and define the quantity

$$\lambda_d \left(\sum_{k=1}^n \xi_{n,k} \rightarrow \xi \right) = \limsup_{n \rightarrow \infty} d \left(\xi, \sum_{k=1}^n \xi_{n,k} \right). \quad (3)$$

Clearly, (3) assigns a positive number to each STA which is 0 if and only if the row-wise sums of the STA are asymptotically normal. Thus this number measures how far a given STA deviates from having an asymptotically normal sequence of row-wise sums.

Now, using the numbers (2) and (3), the first part of Theorem 1.1 leads to the implication

$$\text{Lin}(\{\xi_{n,k}\}) = 0 \Rightarrow \lambda_d \left(\sum_{k=1}^n \xi_{n,k} \rightarrow \xi \right) = 0.$$

Observe that Theorem 1.1 fails to provide any information for the large class of STA's which fail to satisfy Lindeberg's condition, regardless of whether $\text{Lin}(\{\xi_{n,k}\})$ is large or small. Thus the following natural question arises.

Question 1.2. *Suppose that we are given an STA $\{\xi_{n,k}\}$ which is close to satisfying Lindeberg's condition in the sense that $\text{Lin}(\{\xi_{n,k}\})$ is non-zero but small. Is it still possible to conclude that the row-wise sums of $\{\xi_{n,k}\}$ are close to being asymptotically normal in the sense that $\lambda_d(\sum_{k=1}^n \xi_{n,k} \rightarrow \xi)$ is small?*

Let us briefly describe how in the case where d is the Kolmogorov metric

$$K(\eta, \eta') = \sup_{x \in \mathbb{R}} |\mathbb{P}[\eta \leq x] - \mathbb{P}[\eta' \leq x]|,$$

a positive answer to Question 1.2 can be derived from the existing literature.

The following refinement of the sufficiency of Lindeberg's condition in Theorem 1.1 was obtained in terms of the Kolmogorov distance in [O66] and [F68].

Theorem 1.3. *Let ξ be as above. Then there exists a universal constant $C > 0$ such that*

$$K\left(\xi, \sum_{k=1}^n \xi_{n,k}\right) \leq C \left(\sum_{k=1}^n \mathbb{E}[\xi_{n,k}^2; |\xi_{n,k}| > 1] + \sum_{k=1}^n \mathbb{E}[|\xi_{n,k}|^3; |\xi_{n,k}| \leq 1] \right)$$

for all STA's $\{\xi_{n,k}\}$ and all n .

It was shown in [F68] that the constant C in Theorem 1.3 can be taken equal to 6. A proof of Theorem 1.3 based on Stein's method was given in [BH84], and in [CS01], combining Stein's method with Chen's concentration inequality approach, it was established that C can be taken equal to 4.1, the best value known so far up to our knowledge.

We will infer a corollary from Theorem 1.3 which is related to Question 1.2. To this end, we remark that it was pointed out in [L75] that the truncation at 1 in Theorem 1.3 is optimal in the sense that

$$\sum_{k=1}^n \mathbb{E}[\xi_{n,k}^2; |\xi_{n,k}| > 1] + \sum_{k=1}^n \mathbb{E}[|\xi_{n,k}|^3; |\xi_{n,k}| \leq 1]$$

is dominated by

$$\sum_{k=1}^n \mathbb{E}[\xi_{n,k}^2; \xi_{n,k} \in A] + \sum_{k=1}^n \mathbb{E}[|\xi_{n,k}|^3; \xi_{n,k} \in \mathbb{R} \setminus A]$$

for each Borel set $A \subset \mathbb{R}$. Therefore, we easily derive from Theorem 1.3 that

$$K\left(\xi, \sum_{k=1}^n \xi_{n,k}\right) \leq C \left(\sum_{k=1}^n \mathbb{E} [\xi_{n,k}^2; |\xi_{n,k}| > \epsilon] + \epsilon \right)$$

for all $\epsilon > 0$, which, calculating the superior limit of both sides and letting $\epsilon \downarrow 0$, yields

$$\limsup_{n \rightarrow \infty} K\left(\xi, \sum_{k=1}^n \xi_{n,k}\right) \leq C \sup_{\epsilon > 0} \limsup_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{E} [\xi_{n,k}^2; |\xi_{n,k}| > \epsilon].$$

Using the numbers defined in (2) and (3), we now derive the following result as a corollary of Theorem 1.3.

Theorem 1.4. *Let ξ be as above. Then there exists a universal constant $C > 0$ such that*

$$\lambda_K\left(\sum_{k=1}^n \xi_{n,k} \rightarrow \xi\right) \leq C \text{Lin}(\{\xi_{n,k}\})$$

for all STA's $\{\xi_{n,k}\}$.

Remark 1.5. *In [BLV13], combining Stein's method with an asymptotic smoothing technique, it was established that the constant C in Theorem 1.4 can be taken equal to 1 if $\{\xi_{n,k}\}$ satisfies Feller's condition.*

Notice that Theorem 1.4 gives a positive answer to Question 1.2 in the case where $d = K$. It strictly generalizes the sufficiency of Lindeberg's condition in Theorem 1.1, and, contrary to Theorem 1.1, it continues to provide useful information for STA's which have a low Lindeberg index, but fail to satisfy Lindeberg's condition. More precisely, it allows us to conclude that $(\sum_{k=1}^n \xi_{n,k})_n$ is approximately convergent to ξ if $\{\xi_{n,k}\}$ approximately satisfies Lindeberg's condition. Therefore, it seems plausible to refer to Theorem 1.4 as an approximate central limit theorem.

The problem of generalizing Theorem 1.3 to the multivariate setting is hard, and remains open. Notice however that recently, combining a multivariate version of Stein's method, as outlined in e.g. [M09] and [NPR10], with the establishment of an explicit integral representation of a solution to the Stein PDE with a character function as test function, a partial extension of Theorem 1.4 for the Fourier transforms of random vectors has been obtained in [BLV16].

In this paper, we will focus on the following question concerning Theorem 1.4.

Can we widen the scope of applicability of Theorem 1.4 by extending it to other probability metrics d ?

The paper is structured as follows.

A short overview of some important probability metrics is given in section 2.

In section 3, we show that it is possible to apply the techniques used in [BLV13] to a large class of test functions, leading to a general inequality.

The inequality presented in section 3 leads to approximate central limit theorems, similar to Theorem 1.4, for the Wasserstein distance and the parametrized Prokhorov distances. These are given in section 4. An example shows that a result of the same flavor cannot be obtained for the total variation distance.

2. SOME PROBABILITY METRICS

Let $\mathcal{P}(\mathbb{R})$ be the collection of Borel probability measures on \mathbb{R} . Furthermore, let $\mathcal{P}_1(\mathbb{R})$ be the set of all $P \in \mathcal{P}(\mathbb{R})$ with finite absolute first moment, i.e. for which $\int_{-\infty}^{\infty} |x| dP(x) < \infty$.

The Wasserstein distance on $\mathcal{P}_1(\mathbb{R})$, see e.g. [V03], is defined by the formula

$$W(P, Q) = \inf_{\pi} \int_{\mathbb{R} \times \mathbb{R}} d(x, y) d\pi(x, y),$$

where the infimum is taken over all Borel probability measures π on $\mathbb{R} \times \mathbb{R}$ with first marginal P and second marginal Q . Kantorovich duality theory implies that the metric W can also be written as

$$W(P, Q) = \sup_{h \in \mathcal{K}(\mathbb{R})} \left| \int_{\mathbb{R}} h dP - \int_{\mathbb{R}} h dQ \right|, \quad (4)$$

where $\mathcal{K}(\mathbb{R})$ stands for the set of all contractions $h : \mathbb{R} \rightarrow \mathbb{R}$, where h is called a contraction iff $|h(x) - h(y)| \leq |x - y|$ for all $x, y \in \mathbb{R}$. Also, we have

$$W(P, Q) = \int_{-\infty}^{\infty} |F_P(x) - F_Q(x)| dx = \int_0^1 |F_P^{-1}(t) - F_Q^{-1}(t)| dt,$$

with F_P (respectively F_Q) the cumulative distribution function associated with P (respectively Q), and F_P^{-1} (respectively F_Q^{-1}) its generalized inverse.

The topology underlying the Wasserstein distance is slightly stronger than the weak topology. More precisely, for P and $(P_n)_n$ in $\mathcal{P}_1(\mathbb{R})$, it holds that $\lim_{n \rightarrow \infty} W(P, P_n) = 0$ is equivalent with weak convergence of $(P_n)_n$ to P in addition to convergence of $\left(\int_{-\infty}^{\infty} |x| dP_n(x) \right)_n$ to $\int_{-\infty}^{\infty} |x| dP(x)$. Also, the Wasserstein distance is separable and complete, see [B08].

Furthermore, for $\lambda \in \mathbb{R}_0^+$, the (parametrized) Prokhorov distance $\rho_\lambda(P, Q)$ between probability measures P and Q in $\mathcal{P}(\mathbb{R})$ is defined to

be the infimum of all positive numbers $\alpha \in \mathbb{R}_0^+$ for which the inequality

$$P[A] \leq Q[A^{(\lambda\alpha)}] + \alpha,$$

with

$$A^{(\lambda\alpha)} = \left\{ x \in \mathbb{R} \mid \inf_{a \in A} |x - a| \leq \lambda\alpha \right\},$$

holds for every Borel set $A \subset \mathbb{R}$. One easily establishes that

$$\rho_{\lambda_1}(P, Q) \leq \rho_{\lambda_2}(P, Q)$$

whenever $\lambda_2 \leq \lambda_1$. In [B99] it is shown that, for each $\lambda \in \mathbb{R}_0^+$, ρ_λ is a separable and complete metric which metrizes weak convergence of probability measures.

Finally, the total variation distance $d_{TV}(P, Q)$ between probability measures P and Q in $\mathcal{P}(\mathbb{R})$ is defined by the number

$$d_{TV}(P, Q) = \sup_A |P[A] - Q[A]|,$$

the supremum of course taken over all Borel sets $A \subset \mathbb{R}$. One easily verifies that d_{TV} is a complete metric, that, for each $\lambda \in \mathbb{R}_0^+$,

$$\rho_\lambda(P, Q) \leq d_{TV}(P, Q),$$

and that the limit relation

$$\lim_{\lambda \downarrow 0} \rho_\lambda(P, Q) = d_{TV}(P, Q) \tag{5}$$

holds true. Note however that d_{TV} is not separable and that its underlying topology is strictly stronger than the weak topology.

For a general and systematic treatment of the theory of probability metrics, we refer the reader to the excellent expositions [Z83] and [R91].

3. A GENERAL INEQUALITY

Let ξ be as in Section 1 and $h : \mathbb{R} \rightarrow \mathbb{R}$ a continuous map for which $\mathbb{E}|h(\xi)| < \infty$. Then the Stein transform of h is the map $f_h : \mathbb{R} \rightarrow \mathbb{R}$ defined by the formula

$$f_h(x) = e^{x^2/2} \int_{-\infty}^x (h(t) - \mathbb{E}[h(\xi)]) e^{-t^2/2} dt. \tag{6}$$

The crux of Stein's method is that, for any random variable η , we have

$$\mathbb{E}[h(\xi) - h(\eta)] = \mathbb{E}[\eta f_h(\eta) - f'_h(\eta)],$$

and that, in many cases, it is easier to find upper bounds for the derivatives of f_h than for the derivatives of h , see e.g. [BC05] and [CGS11].

We will now establish a general inequality in terms of the Stein transform, which will allow us to extend Theorem 1.4 to many of the above described probability metrics. For the proof, it basically suffices to notice that the techniques developed in [BLV13] can be extended to a very general collection of test functions. For the sake of completeness, we present the proof in Appendix A.

Theorem 3.1. *Let ξ and $\{\xi_{n,k}\}$ be as in Section 1, and let $h : \mathbb{R} \rightarrow \mathbb{R}$ be any continuously differentiable map with a bounded derivative. Then the Stein transform f_h , defined by (6), is twice continuously differentiable, has bounded first and second derivatives, and the inequality*

$$\begin{aligned} & \left| \mathbb{E} \left[h(\xi) - h \left(\sum_{k=1}^n \xi_{n,k} \right) \right] \right| \\ & \leq \frac{1}{2} \|f_h''\|_{\infty} \epsilon + \left(\sup_{x_1, x_2 \in \mathbb{R}} |f_h'(x_1) - f_h'(x_2)| \right) \sum_{k=1}^n \mathbb{E} [\xi_{n,k}^2; |\xi_{n,k}| \geq \epsilon] \\ & \quad + \left(\sup_{x_1, x_2 \in \mathbb{R}} |f_h''(x_1) - f_h''(x_2)| \right) \max_{k=1}^n \mathbb{E} [|\xi_{n,k}|] \end{aligned} \quad (7)$$

holds for all n and all $\epsilon > 0$.

4. APPROXIMATE CENTRAL LIMIT THEOREMS

We will apply Theorem 3.1 to obtain results similar to Theorem 1.4 for the Wasserstein distance (Theorem 4.3) and the parametrized Prokhorov distances (Theorem 4.5). Where needed, we tacitly transport these probability metrics to random variables via their image measures.

The following lemma guarantees that we can capture the Wasserstein distance with continuously differentiable contractions.

Lemma 4.1. *The Wasserstein distance on $\mathcal{P}_1(\mathbb{R})$ is given by*

$$W(P, Q) = \sup_{h \in \mathcal{K}_c(\mathbb{R})} \left| \int h dP - \int h dQ \right|, \quad (8)$$

where $\mathcal{K}_c(\mathbb{R})$ stands for the set of all continuously differentiable contractions $h : \mathbb{R} \rightarrow \mathbb{R}$.

Proof. Let $h : \mathbb{R} \rightarrow \mathbb{R}$ be a contraction and fix $\epsilon > 0$. We will show that there exists a smooth contraction which is closer than ϵ to h for the $\|\cdot\|_{\infty}$ -norm. Once this is established, the lemma will follow from formula (4).

Let

$$\psi_{\epsilon} : \mathbb{R} \rightarrow \mathbb{R}$$

be positive and smooth, with support contained in the interval $[-\epsilon, \epsilon]$, and such that

$$\int_{\mathbb{R}} \psi_{\epsilon}(y) dy = 1.$$

Put

$$h_{\epsilon}(x) = (h \star \psi_{\epsilon})(x) = \int_{\mathbb{R}} h(x - y) \psi_{\epsilon}(y) dy = \int_{\mathbb{R}} \psi_{\epsilon}(x - y) h(y) dy.$$

Then h_{ϵ} is smooth. Furthermore, for $x_1, x_2 \in \mathbb{R}$,

$$|h_{\epsilon}(x_1) - h_{\epsilon}(x_2)|$$

$$\begin{aligned}
&= \left| \int_{\mathbb{R}} h(x_1 - y) \psi_{\epsilon}(y) dy - \int_{\mathbb{R}} h(x_2 - y) \psi_{\epsilon}(y) dy \right| \\
&\leq \int_{\mathbb{R}} |h(x_1 - y) - h(x_2 - y)| \psi_{\epsilon}(y) dy,
\end{aligned}$$

which is, h being a contraction, bounded by $\int_{\mathbb{R}} \psi_{\epsilon}(y) dy = 1$, and we infer that h_{ϵ} is also a contraction. Finally, for $x \in \mathbb{R}$,

$$\begin{aligned}
&|h(x) - h_{\epsilon}(x)| \\
&= \left| \int_{\mathbb{R}} (h(x) - h(x - y)) \psi_{\epsilon}(y) dy \right| \\
&= \left| \int_{-\epsilon}^{\epsilon} (h(x) - h(x - y)) \psi_{\epsilon}(y) dy \right|, \tag{9}
\end{aligned}$$

the last equality following from the fact that the support of ψ_{ϵ} is contained in $[-\epsilon, \epsilon]$. Now, h being a contraction, it follows that the expression in (9) is bounded by ϵ , whence

$$\|h - h_{\epsilon}\|_{\infty} < \epsilon.$$

This concludes the proof. \square

The following lemma belongs to the basics of Stein's method, see e.g. [BC05], p.10-11.

Lemma 4.2. *Let h and f_h be as in Theorem 3.1. Then*

$$\|f'_h\|_{\infty} \leq 4\|h'\|_{\infty} \tag{10}$$

and

$$\|f'_h\|_{\infty} \leq 2\|\mathbb{E}[h(\xi)] - h\|_{\infty} \tag{11}$$

and

$$\|f''_h\|_{\infty} \leq 2\|h'\|_{\infty}. \tag{12}$$

Theorem 4.3. *Let ξ be as in Section 1. Then there exists a universal constant $C_W > 0$ such that*

$$\lambda_W \left(\sum_{k=1}^n \xi_{n,k} \rightarrow \xi \right) \leq C_W \text{Lin}(\{\xi_{n,k}\})$$

for all STA's $\{\xi_{n,k}\}$ which satisfy Feller's condition (1). Moreover, C_W can be taken equal to 8.

Proof. Let $h : \mathbb{R} \rightarrow \mathbb{R}$ be a continuously differentiable contraction. Then

$$\sup_{x_1, x_2 \in \mathbb{R}} |f'_h(x_1) - f'_h(x_2)| \leq 2\|f'_h\|_{\infty} \leq 8\|h'\|_{\infty} \leq 8, \tag{13}$$

by (10), and

$$\sup_{x_1, x_2 \in \mathbb{R}} |f''_h(x_1) - f''_h(x_2)| \leq 2\|f''_h\|_{\infty} \leq 4\|h'\|_{\infty} \leq 4, \tag{14}$$

by (12). Furthermore, combining (7) with (12), (13), and (14), yields

$$\begin{aligned} & \left| \mathbb{E} \left[h(\xi) - h \left(\sum_{k=1}^n \xi_{n,k} \right) \right] \right| \\ & \leq \epsilon + 8 \sum_{k=1}^n \mathbb{E} [\xi_{n,k}^2; |\xi_{n,k}| \geq \epsilon] + 4 \max_{k=1}^n \mathbb{E} [|\xi_{n,k}|] \end{aligned} \quad (15)$$

for all STA's $\{\xi_{n,k}\}$, all n , and all $\epsilon > 0$. Finally, assuming that $\{\xi_{n,k}\}$ satisfies Feller's condition, taking the supremum over all $h \in \mathcal{K}_c(\mathbb{R})$, calculating the superior limits, and letting $\epsilon \downarrow 0$, we see that that (8) and (15) lead to the desired result. \square

Lemma 4.4 reveals that we can capture all parametrized Prokhorov distances by one collection of smooth test functions. It can be derived indirectly from [BLV11] (Section 2, Lemma 2.2), where the so-called weak approach structure on the set of probability measures on a separable metric space was studied, see also [L15]. As it is a crucial step to obtain an approximate central limit theorem for the parametrized Prokhorov distances, we will present a direct proof here.

Lemma 4.4. *Let \mathcal{H} be the collection of continuously differentiable maps $h : \mathbb{R} \rightarrow [0, 1]$ with a bounded derivative. Then, for P and $(P_n)_n$ in $\mathcal{P}(\mathbb{R})$,*

$$\sup_{\lambda > 0} \limsup_{n \rightarrow \infty} \rho_\lambda(P, P_n) = \sup_{h \in \mathcal{H}} \limsup_{n \rightarrow \infty} \left| \int_{\mathbb{R}} h dP - \int_{\mathbb{R}} h dP_n \right|. \quad (16)$$

Proof. First suppose that, for $\gamma > 0$,

$$\sup_{h \in \mathcal{H}} \limsup_{n \rightarrow \infty} \left| \int_{\mathbb{R}} h dP - \int_{\mathbb{R}} h dP_n \right| < \gamma. \quad (17)$$

Now fix $\epsilon > 0$ and $\lambda > 0$, and choose real numbers

$$a_1 < a_2 < \dots < a_{N-1} < a_N$$

such that

$$\forall k \in \{1, \dots, N-1\} : a_{k+1} - a_k < \lambda\gamma/2 \quad (18)$$

and

$$P[\mathbb{R} \setminus [a_1, a_N]] < \epsilon. \quad (19)$$

For each subset $K \subset \{1, \dots, N-1\}$, choose $h_K \in \mathcal{H}$ such that

$$\forall x \in \cup_{k \in K} [a_k, a_{k+1}] : h_K(x) = 1 \quad (20)$$

and

$$\forall x \in \mathbb{R} \setminus (\cup_{k \in K} [a_k, a_{k+1}])^{(\lambda\gamma/2)} : h_K(x) = 0. \quad (21)$$

By (17), for each $K \subset \{1, \dots, N-1\}$, there exists n_K such that for all $n \geq n_K$

$$\left| \int_{\mathbb{R}} h_K dP - \int_{\mathbb{R}} h_K dP_n \right| < \gamma. \quad (22)$$

Let $n_0 = \max_{K \subset \{1, \dots, N-1\}} n_K$, and take $n \geq n_0$ and a Borel set $A \subset \mathbb{R}$. Put

$$K_0 = \{k \in \{1, \dots, N-1\} : A \cap [a_k, a_{k+1}] \neq \emptyset\}. \quad (23)$$

Then

$$P[A] \leq P[\cup_{k \in K_0} [a_k, a_{k+1}]] + P[\mathbb{R} \setminus [a_1, a_N]],$$

which, by (19) and (20),

$$\leq \int_{\mathbb{R}} h_{K_0} dP + \epsilon,$$

which, by (22),

$$< \int_{\mathbb{R}} h_{K_0} dP_n + \gamma + \epsilon,$$

which, by (21),

$$\leq P_n \left[(\cup_{k \in K_0} [a_k, a_{k+1}])^{(\lambda\gamma/2)} \right] + \gamma + \epsilon,$$

which, by (18) and (23),

$$\leq P_n [A^{(\lambda\gamma)}] + \gamma + \epsilon.$$

By definition of the parametrized Prokhorov metric, we conclude that

$$\rho_\lambda(P, P_n) \leq \gamma + \epsilon.$$

Since n , λ , and ϵ were arbitrary, we infer that

$$\sup_{\lambda > 0} \limsup_n \rho_\lambda(P, P_n) \leq \gamma.$$

Finally, since γ was arbitrarily taken such that (17) holds, we conclude that the left-hand side of (16) is dominated by the right-hand side of (16).

For the reverse inequality, suppose that, for $\gamma > 0$,

$$\sup_{\lambda > 0} \limsup_{n \rightarrow \infty} \rho_\lambda(P, P_n) < \gamma. \quad (24)$$

Now fix $h \in \mathcal{H}$ and $\epsilon > 0$. Notice that, by the monotone convergence theorem, $\int_0^1 P[\{h \geq t\}^{(\lambda\gamma)}] dt \rightarrow \int_0^1 P[h \geq t] dt$ as $\lambda \downarrow 0$, whence we find $\lambda_0 > 0$ such that

$$\int_0^1 P[\{h \geq t\}^{(\lambda_0\gamma)}] dt \leq \int_0^1 P[h \geq t] dt + \epsilon. \quad (25)$$

By (24), and using symmetry of ρ_λ , we find n_0 such that for each $n \geq n_0$ and each Borel set $A \subset \mathbb{R}$

$$P_n[A] \leq P[A^{(\lambda_0\gamma)}] + \gamma. \quad (26)$$

Fix $n \geq n_0$. By the layer cake representation,

$$\int h dP_n = \int_0^1 P_n[h \geq t] dt,$$

which, by (26),

$$\leq \int_0^1 P[\{h \geq t\}^{(\lambda_0 \gamma)}] dt + \gamma,$$

which, by (25),

$$\leq \int_0^1 P[h \geq t] dt + \gamma + \epsilon,$$

which, again by the layer cake representation,

$$= \int h dP + \gamma + \epsilon.$$

As before, by the arbitrariness of n , h , ϵ , and γ , we conclude that

$$\limsup_{n \rightarrow \infty} \left(\int h dP_n - \int h dP \right) \leq \sup_{\lambda > 0} \limsup_{n \rightarrow \infty} \rho_\lambda(P, P_n). \quad (27)$$

Finally, taking into account that $h \in \mathcal{H}$ if and only if $1 - h \in \mathcal{H}$, (27) learns that the right-hand side of (16) is dominated by the left-hand side of (16), which finishes the proof. \square

Theorem 4.5. *Let ξ be as in Section 1. Then there exists a universal constant $C_P > 0$ such that*

$$\lambda_{\rho_\lambda} \left(\sum_{k=1}^n \xi_{n,k} \rightarrow \xi \right) \leq C_P \text{Lin}(\{\xi_{n,k}\})$$

for all $\lambda > 0$ and all STA's $\{\xi_{n,k}\}$ which satisfy Feller's condition (1). Moreover, C_P can be taken equal to 4.

Proof. Let $h : \mathbb{R} \rightarrow [0, 1]$ be a continuously differentiable map with a bounded derivative. Then

$$\sup_{x_1, x_2 \in \mathbb{R}} |f'_h(x_1) - f'_h(x_2)| \leq 2\|f'_h\|_\infty \leq 4\|\mathbb{E}[h(\xi)] - h\|_\infty \leq 4, \quad (28)$$

by (11), and

$$\sup_{x_1, x_2 \in \mathbb{R}} |f''_h(x_1) - f''_h(x_2)| \leq 2\|f''_h\|_\infty \leq 4\|h'\|_\infty, \quad (29)$$

by (12). Furthermore, combining (7) with (12), (28), and (29), yields

$$\begin{aligned} & \left| \mathbb{E} \left[h(\xi) - h \left(\sum_{k=1}^n \xi_{n,k} \right) \right] \right| \\ & \leq \|h'\|_\infty \epsilon + 4 \sum_{k=1}^n \mathbb{E} [\xi_{n,k}^2; |\xi_{n,k}| \geq \epsilon] + 4\|h'\|_\infty \max_{k=1}^n \mathbb{E} [|\xi_{n,k}|] \end{aligned} \quad (30)$$

for all STA's $\{\xi_{n,k}\}$, all n , and all $\epsilon > 0$. Finally, assuming that $\{\xi_{n,k}\}$ satisfies Feller's condition, calculating the superior limits, and letting $\epsilon \downarrow 0$, we see that that (16) and (30) lead to the desired result. \square

Notice the remarkable fact that the constant C_P in Theorem 4.5 does not depend on the parameter λ . This, in light of relation (5), suggests that an approximate central limit theorem in the spirit of Theorem 1.4 for the total variation distance d_{TV} might be derived from Theorem 4.5. However, the following example shows that this is not the case.

Example 4.6. *Let ξ and $\{\xi_{n,k}\}$ be as in Section 1, and assume that $\{\xi_{n,k}\}$ consists of discrete random variables and satisfies Lindeberg's condition. Then*

$$\text{Lin}(\{\xi_{n,k}\}) = 0$$

and, each $\sum_{k=1}^n \xi_{n,k}$ also being discrete,

$$d_{TV}\left(\xi, \sum_{k=1}^n \xi_{n,k}\right) = 1.$$

We conclude that there does not exist a constant $C > 0$ such that

$$\lambda_{d_{TV}}\left(\sum_{k=1}^n \xi_{n,k} \rightarrow \xi\right) \leq C \text{Lin}(\{\xi_{n,k}\})$$

for all STA's $\{\xi_{n,k}\}$ satisfying Feller's condition.

We summarize the information obtained in Theorem 1.4 and Remark 1.5, Theorem 4.3, and Theorem 4.5, in the following result. We put

$$\lambda_P\left(\sum_{k=1}^n \xi_{n,k} \rightarrow \xi\right) = \sup_{\lambda \in \mathbb{R}_0^+} \lambda_{\rho_\lambda}\left(\sum_{k=1}^n \xi_{n,k} \rightarrow \xi\right).$$

Theorem 4.7. *Let ξ be as in Section 1. Then, for each $\delta \in \{K, W, P\}$, there exists a universal constant $C_\delta > 0$ such that*

$$\lambda_\delta\left(\sum_{k=1}^n \xi_{n,k} \rightarrow \xi\right) \leq C_\delta \text{Lin}(\{\xi_{n,k}\})$$

for all STA's $\{\xi_{n,k}\}$ satisfying Feller's condition (1). Moreover, C_K can be taken equal to 1, C_W equal to 8, and C_P equal to 4.

APPENDIX A: PROOF OF THEOREM 3.1

We follow [BLV13], Section 2. We keep a continuously differentiable $h : \mathbb{R} \rightarrow [0, 1]$, with bounded derivative, fixed, and let f_h be its Stein transform defined by (6). Also, we put

$$\sigma_{n,k}^2 = \mathbb{E}[\xi_{n,k}^2].$$

The following lemma is easily verified. It can be found in e.g. [BC05] (p.10-11).

Lemma 1. *f_h is twice continuously differentiable, has bounded first and second derivatives, and*

$$\mathbb{E}[h(\xi)] - h(x) = x f_h(x) - f_h'(x). \quad (31)$$

The following lemma can be found in [BLV13] (Lemma 2.4). We give the proof for completeness.

Lemma 2. *Put*

$$\delta_{n,k} = f_h \left(\sum_{i \neq k} \xi_{n,i} + \xi_{n,k} \right) - f_h \left(\sum_{i \neq k} \xi_{n,i} \right) - \xi_{n,k} f'_h \left(\sum_{i \neq k} \xi_{n,i} \right)$$

and

$$\epsilon_{n,k} = f'_h \left(\sum_{i \neq k} \xi_{n,i} + \xi_{n,k} \right) - f'_h \left(\sum_{i \neq k} \xi_{n,i} \right) - \xi_{n,k} f''_h \left(\sum_{i \neq k} \xi_{n,i} \right).$$

Then

$$\begin{aligned} & \mathbb{E} \left[\left(\sum_{k=1}^n \xi_{n,k} \right) f_h \left(\sum_{k=1}^n \xi_{n,k} \right) - f'_h \left(\sum_{k=1}^n \xi_{n,k} \right) \right] \\ &= \sum_{k=1}^n \mathbb{E} [\xi_{n,k} \delta_{n,k}] - \sum_{k=1}^n \sigma_{n,k}^2 \mathbb{E} [\epsilon_{n,k}]. \end{aligned} \quad (32)$$

Proof. Recalling that $\xi_{n,k}$ and $\sum_{i \neq k} \xi_{n,i}$ are independent, $\mathbb{E} [\xi_{n,k}] = 0$, and $\sum_{k=1}^n \sigma_{n,k}^2 = 1$, we get

$$\begin{aligned} & \sum_{k=1}^n \mathbb{E} [\xi_{n,k} \delta_{n,k}] - \sum_{k=1}^n \sigma_{n,k}^2 \mathbb{E} [\epsilon_{n,k}] \\ &= \sum_{k=1}^n \mathbb{E} \left[\xi_{n,k} f_h \left(\sum_{k=1}^n \xi_{n,k} \right) \right] - \mathbb{E} \left[\xi_{n,k} f_h \left(\sum_{i \neq k} \xi_{n,i} \right) \right] \\ & \quad - \sum_{k=1}^n \mathbb{E} \left[\xi_{n,k}^2 f'_h \left(\sum_{i \neq k} \xi_{n,i} \right) \right] - \sum_{k=1}^n \sigma_{n,k}^2 \mathbb{E} \left[f'_h \left(\sum_{k=1}^n \xi_{n,k} \right) \right] \\ & \quad + \sum_{k=1}^n \mathbb{E} [\xi_{n,k}^2] \mathbb{E} \left[f'_h \left(\sum_{i \neq k} \xi_{n,i} \right) \right] + \sum_{k=1}^n \sigma_{n,k}^2 \mathbb{E} \left[\xi_{n,k} f''_h \left(\sum_{i \neq k} \xi_{n,i} \right) \right]. \end{aligned}$$

The last expression further reduces to

$$\begin{aligned} & \mathbb{E} \left[\left(\sum_{k=1}^n \xi_{n,k} \right) f_h \left(\sum_{k=1}^n \xi_{n,k} \right) \right] - \mathbb{E} [\xi_{n,k}] \mathbb{E} \left[f_h \left(\sum_{i \neq k} \xi_{n,i} \right) \right] \\ & \quad - \sum_{k=1}^n \mathbb{E} \left[\xi_{n,k}^2 f'_h \left(\sum_{i \neq k} \xi_{n,i} \right) \right] - \mathbb{E} \left[f'_h \left(\sum_{k=1}^n \xi_{n,k} \right) \right] \\ & \quad + \sum_{k=1}^n \mathbb{E} [\xi_{n,k}^2] \mathbb{E} \left[f'_h \left(\sum_{i \neq k} \xi_{n,i} \right) \right] + \sum_{k=1}^n \sigma_{n,k}^2 \mathbb{E} [\xi_{n,k}] \mathbb{E} \left[f''_h \left(\sum_{i \neq k} \xi_{n,i} \right) \right], \end{aligned}$$

which is easily seen to equal

$$\mathbb{E} \left[\left(\sum_{k=1}^n \xi_{n,k} \right) f_h \left(\sum_{k=1}^n \xi_{n,k} \right) - f'_h \left(\sum_{k=1}^n \xi_{n,k} \right) \right].$$

This finishes the proof. \square

The following lemma is an application of Taylor's theorem.

Lemma 3. *For any $a, x \in \mathbb{R}$,*

$$\begin{aligned} & |f_h(a+x) - f_h(a) - f'_h(a)x| \\ & \leq \min \left\{ \left(\sup_{x_1, x_2 \in \mathbb{R}} |f'_h(x_1) - f'_h(x_2)| \right) |x|, \frac{1}{2} \|f''_h\|_\infty x^2 \right\}. \end{aligned} \quad (33)$$

We are now in a position to present a proof of Theorem 3.1.

Proof of Theorem 3.1. For n and $\epsilon > 0$, we have, by (31), (32), and (33),

$$\begin{aligned} & \left| \mathbb{E} \left[h(\xi) - h \left(\sum_{k=1}^n \xi_{n,k} \right) \right] \right| \\ &= \left| \mathbb{E} \left[\left(\sum_{k=1}^n \xi_{n,k} \right) f_h \left(\sum_{k=1}^n \xi_{n,k} \right) - f'_h \left(\sum_{k=1}^n \xi_{n,k} \right) \right] \right| \\ &\leq \sum_{k=1}^n \mathbb{E} [|\xi_{n,k} \delta_{n,k}|] + \sum_{k=1}^n \sigma_{n,k}^2 \mathbb{E} [|\epsilon_{n,k}|] \\ &\leq \frac{1}{2} \|f''_h\|_\infty \sum_{k=1}^n \mathbb{E} [|\xi_{n,k}|^3; |\xi_{n,k}| < \epsilon] \\ &\quad + \left(\sup_{x_1, x_2 \in \mathbb{R}} |f'_h(x_1) - f'_h(x_2)| \right) \sum_{k=1}^n \mathbb{E} [|\xi_{n,k}|^2; |\xi_{n,k}| \geq \epsilon] \\ &\quad + \left(\sup_{x_1, x_2 \in \mathbb{R}} |f''_h(x_1) - f''_h(x_2)| \right) \sum_{k=1}^n \sigma_{n,k}^2 \mathbb{E} [|\xi_{n,k}|], \end{aligned}$$

which proves the desired result since $\sum_{k=1}^n \sigma_{n,k}^2 = 1$.

REFERENCES

- [BC05] Barbour, A.D.; Chen, L.H.Y. *An introduction to Stein's method* Singapore University Press, Singapore; World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005.
- [BH84] Barbour, A. D.; Hall, P. *Stein's method and the Berry-Esseen theorem*. Austral. J. Statist. 26 (1984), no. 1, 8–15.
- [BLV11] Berckmoes, B.; Lowen, R.; Van Casteren, J. Distances on probability measures and random variables. J. Math. Anal. Appl. 374 (2011), no. 2, 412–428.

- [BLV13] Berckmoes, B.; Lowen, R.; Van Casteren, J. *An isometric study of the Lindeberg-Feller central limit theorem via Stein's method*. J. Math. Anal. Appl. 405 (2013), no. 2, 484–498.
- [BLV16] Berckmoes B.; Lowen R.; Van Casteren J. *Stein's Method and a Quantitative Lindeberg CLT for the Fourier Transforms of Random Vectors* J. Math. Anal. Appl. 433 (2016), no. 2, 1441–1458.
- [B99] Billingsley, P. *Convergence of probability measures* Second edition. Wiley Series in Probability and Statistics: Probability and Statistics. A Wiley-Interscience Publication. John Wiley and Sons, Inc., New York, 1999.
- [B08] Bolley, F. *Separability and completeness for the Wasserstein distance*. Séminaire de probabilités XLI, 371–377, Lecture Notes in Math., 1934, Springer, Berlin, 2008.
- [CGS11] Chen, L. H. Y.; Goldstein, L.; Shao, Q-M. *Normal approximation by Stein's method*. Probability and its Applications (New York). Springer, Heidelberg, 2011.
- [CS01] Chen, L. H. Y.; Shao, Q-M. *A non-uniform Berry-Esseen bound via Stein's method*. Probab. Theory Related Fields 120 (2001), no. 2, 236–254.
- [O66] Osipov, L. V. *A refinement of Lindberg's theorem*. (Russian) Teor. Veroyatnost. i Primenen. 11 1966 339–342.
- [F68] Feller, W. *On the Berry-Esseen theorem*. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 10 1968 261–268.
- [L75] Loh, W.Y. *On the normal approximation for sums of mixing random variables*. Master Thesis, Department of Mathematics, University of Singapore (1975)
- [L15] Lowen, R. *Index analysis. Approach theory at work*. Springer Monographs in Mathematics. Springer, London, 2015.
- [M09] Meckes, E. *On Stein's method for multivariate normal approximation*. High dimensional probability V: the Luminy volume, 153–178, Inst. Math. Stat. Collect., 5, Inst. Math. Statist., Beachwood, OH, 2009.
- [NPR10] Nourdin, I.; Peccati, G.; Réveillac, A. *Multivariate normal approximation using Stein's method and Malliavin calculus*. Ann. Inst. Henri Poincaré Probab. Stat. 46 (2010), no. 1, 45–58.
- [R91] Rachev, S. T. *Probability metrics and the stability of stochastic models*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 1991.
- [V03] Villani, C. *Topics in optimal transportation*. Graduate Studies in Mathematics, 58. American Mathematical Society, Providence, RI, 2003.
- [Z83] Zolotarev, V. M. *Probability metrics*. (Russian) Teor. Veroyatnost. i Primenen. 28 (1983), no. 2, 264–287.